

Genetic algorithms tuned expert model for detection of epileptic seizures from EEG signatures

Rohtash Dhiman*, J.S. Saini, Priyanka

Electrical Engineering Deptt., Deenbandhu Chhotu Ram University of Science & Technology, Murthal, District Sonapat, Haryana 131039, India

ARTICLE INFO

Article history:

Received 29 September 2012

Received in revised form

10 November 2013

Accepted 16 January 2014

Available online 6 February 2014

Keywords:

Discrete wavelet transform (DWT)

Genetic algorithm (GA)

Support vector machine (SVM)

Electroencephalogram (EEG)

Field programmable gate arrays (FPGA)

ABSTRACT

The uncontrolled firing of neurons in brain leads to epileptic seizures in the patients. A novel scheme to detect epileptic seizures from back ground electroencephalogram (EEG) is proposed in this paper. This scheme is based on discrete wavelet packet transform with energy, entropy, kurtosis, skewness, mean, median and standard deviation as the properties for creating features of signals for classification. Optimal features are selected using genetic algorithm (GA) with support vector machine as a classifier for creating objective function values for the GA. Clinical EEG data from epileptic and normal subjects are used in the experiment. The knowledge of neurologist (medical expert) is utilized to train the system. To evaluate the efficacy of the proposed scheme, a 10 fold cross-validation is implemented, and the detection rate is found 100% accurate with 100% of sensitivity and specificity for the data under consideration. The proposed GA-SVM scheme is a novel technique using a hybrid approach with wavelet packet decomposition, support vector machine and GA. It is novel in terms of selection of features sub set, use of SVM classifier as objective function for GA and improved classification rate. The proposed model can be used in the developing and the third world countries where the medical facilities are in acute shortage and qualified neurologists are not available. This system can be helpful in assisting the neurologists in terms of automated observation and saving valuable human expert time.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Electroencephalogram (EEG) was first measured in humans by German psychiatrist Hans Berger [6] in 1929. It gives an insight into the neural activity in brain and is a non-invasive method to study cognitive processes and the physiology of the brain. It gives very useful information related to brain dynamics and different physiological and psychological aspects of the brain. Epilepsy is a chronic disorder characterized by recurrent seizures, which may vary from a brief lapse of attention or muscle jerks, to severe and prolonged convulsions in the body [47]. The seizures are caused by sudden, usually brief, excessive electrical discharges in a group of brain cells (neurons) (Cited in WHO: <http://www.who.int/topics/epilepsy/en/>). More than 1% (50 million people) of the world's population is affected by epilepsy [44]. Eighty percent of the epileptic seizure activity can be controlled or treated effectively, if properly detected and diagnosed (cited in WHO: <http://www.who.int/topics/epilepsy/en/>) [47]. The acute shortage of medical facilities in the developing and third world

countries is the major cause of deaths. Even if the basic facilities of medical help may be present in some parts of these countries, yet trained neurologists are present only in super specialty hospitals in big cities. A neurologist inspects the EEG recordings over time to assess and diagnose various ailments related to brain and thereby decides the course of treatment and evaluates the pre-surgical conditions of patient if surgery is required. In rural and remote areas, some form of automated system of diagnosis of severe illness like epilepsy is required which can be operated by skilled or semi-skilled medical persons. Other benefits of designing an automated expert system are to help the neurologist to effectively assess the condition, to reduce the requirement for storage of data and to save time of manual inspection.

Many approaches towards automatic detection of this disease are reported in literature [1,3,16,18–20,23,31,32,35,37,52]. In this paper, a novel scheme is proposed, which uses the knowledge of a trained neurologist about the epileptic and non-epileptic subjects for training of the system, and once trained, it can detect epileptic seizures. The performance of an epileptic seizure detection system depends on features extraction and features selection. In this paper, the use of discrete wavelet packet decomposition with energy, entropy, kurtosis, skewness, mean, median and standard deviation, as properties for creating the feature set is proposed. A genetic

* Corresponding author. Tel.: +91 9991425740.

E-mail addresses: rohtash.k@gmail.com, rohtash.ee@dcrustm.org (R. Dhiman).

algorithm with adaptive probability of mutation is used as the optimization algorithm for selecting optimal feature set to present to the classifier so as to maximize the classification accuracy. Support vector machine is used as a classifier for each set of values stored in each chromosome and the classification rate is chosen as the objective function value for the GA. The data of epileptic and non-epileptic subjects are taken from Andrzejak website as given in [40]. The proposed GA-SVM technique outperforms the results of recent approaches reported in [20,37,52]. In the present approach, optimal selection of feature sub set is employed, so it outperforms the approaches of [4] in terms of feature set and also in terms of correct classification rate. The proposed approach outperforms the reported techniques in that it gives 100% correct classification rate whereas the earlier approaches reported the same between 73% [37] and 99.33% [47]. The paper is organized in six sections starting with introductory section followed by literature survey and state of the art in Section 2. Data and methods are given in Section 3 followed by optimization of features selection in Section 4. Section 5 deals with experimental results and Section 6 presents conclusions and future directions.

2. Literature survey and state of the art

Richard Canton, an English physician, discovered electrical currents in the brain in 1875. Electroencephalogram (EEG) was first recorded [6] in 1929, and paved the way for studying electrical signals related to the cognitive and brain activities. EEG is a non-invasive method giving a lot of information related to the different physiological states and complex dynamical behaviour of the brain. EEG recorded over a long time span is very important for monitoring incidental disorders like epileptic seizures which are not permanently present in the recordings. EEG data for long periods of time is visually inspected by trained neurologists for detecting epileptic seizures and pre-surgical evaluation of epilepsy patients, and used for clinical diagnosis and possible treatment plans. This is a time consuming and very costly process. Efforts for predicting the epileptic seizures were first time reported in 1975 [43]; however, further attention was paid by researchers and clinical neurologists during 1990s. A recent approach (2012) presented in [52] employs a tunable support vector machine classifier for epileptic seizure detection, and gives correct classification rate of 98.72%. Tapan et al. [47] recently proposed a scheme based on discrete wavelet transform (DWT) and energy estimation at each node of the decomposition tree followed by application of probabilistic neural network (PNN) for classification, giving classification rate of 99.33%. Recently Zandi et al. reported predicting of epileptic seizures in scalp EEG based on a Variational Bayesian Gaussian mixture model of zero-crossing intervals [1]. An 8-channel scalable EEG acquisition SoC (System-on-Chip) with patient-specific seizure classification and recording processor is presented in [23]. An EEG classification system based on Wavelet-CSP for hand movements is presented in [36]. Neonatal seizure detection is addressed in [32]. The canonical correlation is used in [16], while in [35], detection of epileptic spikes by magnetoencephalography and electroencephalography (EEG) is explored. A data driven approach for identification of pre-seizure states in epilepsy is presented in [19]. Recurrent neural network based prediction of epileptic seizures in intra- and extra-cranial EEG is reported in [3].

A comparative study based on the field of seizure prediction using statistical validation of event predictors is given in [18]. A comparative analysis of feature selection from EEG, using mutual information and support vector machine is presented in [7]. Prediction of epileptic seizures using accumulated energy in a multi-resolution framework is presented in [45]. Some approaches for seizure detection like Phase space topography and

the Lyapunov exponent of electrocorticograms [30], computer analysis of EEG signals with parametric models [2], artificial neural networks [38], non-linear time series analysis [24,26,33], recurrence quantification analysis [51], synchronization in intracranial electroencephalogram recordings [15] have been reported in literature. Nonlinear dynamics and chaos are used in [24,26,30,51], phase coherence in [32], and time domain analysis in [4,5], for seizure detection based on EEG. Various soft computing approaches for scalp EEG processing are explained in [25]. EEG signals are used for classification of BMD and ADHD patients [27]. Seizures anticipation in human neocortical partial epilepsy is addressed in [49]. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model is presented in [48]. The methods reported in [37] by Ocak use genetic algorithms and wavelet transforms but only entropy at different levels of decomposition is taken for preparation of feature sets and a correct classification rate of 73% was achieved. The same authors attempted discrete wavelet transform and approximate entropy leading to an increased classification rate of 96% [20]. In [13], the use of mixture of experts (ME) network structure to guide model selection for classification of electroencephalogram (EEG) signals is illustrated and gives classification rate of 93.17%. In this paper, a novel scheme is proposed based on discrete wavelet packet decomposition of the signals from epileptic and non-epileptic subjects (100 segments each) and then computing the energy and other statistical and non-statistical properties at each node of decomposition tree to create the feature set for classification. The feature set thus created is vast enough, so an optimal selection of the feature sub-set is done by using a GA. This technique outperforms the results of recent approaches reported in [20,37,52]. As there is no feature sub-set selected in the approaches reported in [47], but in the present approach optimal selection of feature sub set is employed, so it outperforms the approaches of [47] in terms of feature set and also in terms of correct classification rate. The proposed approach outperforms the reported techniques in that it gives 100% correct classification rate where as the earlier approaches reported the same between 73% [37] and 99.33% [47].

3. Data and methods

3.1. Data selection

The EEG data used for this work is taken from University of Bonn, Germany, as given in [40] by Andrzejak et al. This data [40],

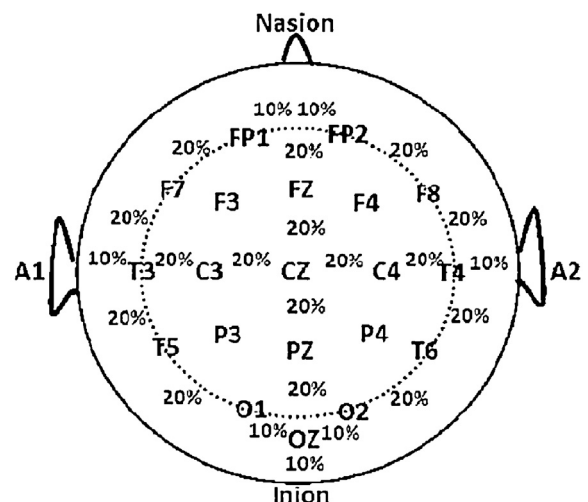


Fig. 1. 10/20 International system of electrode placement.

comprising of five sets (denoted A–E), each containing 100 EEG segments of 23.6-s duration, are used for the study. Sets A and B consist of segments taken from surface EEG recordings that are carried out on healthy volunteers using a standardized electrode placement scheme (10/20 International system of electrode placement for non-invasive recording of EEG from human scalp) as shown in Fig. 1. Volunteers were relaxed in an awakened state with eyes open (A) and eyes closed (B), respectively. While sets C and D contain only activity measured during seizure free intervals, set E contains only seizure activity. Segments are selected from all recording sites exhibiting ictal activity [40]. We take set E and set A for our experiments, as set E contains the ictal state data, so these are confirmed cases of epilepsy and set A pertains to confirmed subjects having no epilepsy. Knowledge of medical expert is available in terms of the data sets already defined to be containing data from epileptic or non-epileptic subjects. Part of this data (i.e., 50%) is used for training and 50% for testing during execution of GA-SVM and a 10-fold cross validation is used for the final optimized feature sub set given by GA-SVM, to illustrate the efficacy of the technique.

3.2. Wavelet packet decomposition and feature extraction

Wavelet packet analysis is used for extracting features of non-stationary signals, and as such is appropriate for extracting features from EEG signals which are non-stationary in nature. Wavelet transform gives time-frequency representation of a signal. Wavelet packet analysis is a generalized form of the DWT, wherein a signal is split into approximation and detailed coefficients. The approximation is again split into a second-level approximation and detailed coefficients, and the process is repeated. There are $k+1$ possible ways to encode the signal for the k -level decomposition. The signal is passed through a series of low-pass and high-pass filters called quadrature mirror filters. In wavelet packet analysis, the approximation as well as detailed coefficients can be split like a complete binary tree structure. This way, of recursively applying decomposition process in both low and high frequency sub-bands to generate next level in the binary tree, yields more than 2^{2k-1} different ways to encode the signal. For each level k , there are 2^{k-1} number of nodes available after wavelet packet decomposition. Fig. 2 shows the wavelet packet decomposition tree obtained for 8th level of decomposition and example data for node (8, 64) means data at 8th level and 64th coefficient of decomposition. Shannon entropy-based criterion is used to select suitable decomposition of the signal. Eighth level decomposition coefficients are used to extract the features. Statistical and non-statistical properties such as energy, mean, median, standard deviation, kurtosis, skewness and Shannon entropy of the decomposition coefficients are used to create the complete feature set. All algorithms are developed in the MATLAB environment with Wavelet and Bioinformatics Toolboxes. The coefficients C_{ik} at 8th decomposition level are used to calculate features by using following standard equations [47]:

$$\text{Energy, } ED_i = \sum_{k=1}^N |C_{ik}|^2 \quad i = 1, 2, \dots, l \quad (1)$$

$$\text{Standard deviation, } \sigma_i = \frac{1}{N-1} \sum_{k=1}^N (C_{ik} - \mu_i)^2, \quad i = 1, 2, \dots, l \quad (2)$$

$$\text{Mean, } \mu_i = \frac{1}{N} \sum_{k=1}^N C_{ik}, \quad i = 1, 2, \dots, l \quad (3)$$

$$\text{Kurtosis, } KUR_i = \frac{\varepsilon(C_{ik} - \mu_i)^3}{\sigma_i^3}, \quad i = 1, 2, \dots, l \quad (4)$$

$$\text{Skewness, } SK_i = \frac{\varepsilon(C_{ik} - \mu_i)^3}{\sigma_i^3}, \quad i = 1, 2, \dots, l \quad (5)$$

$$\text{Entropy, } EN_i = \sum_{k=1}^N C_{ik}^2 \log(C_{ik}^2), \quad i = 1, 2, \dots, l \quad (6)$$

$$\text{Median, } med_i = \sum_{k=1}^N \text{median}(C_{ik}), \quad i = 1, 2, \dots, l \quad (7)$$

The coefficient C_{ik} is the decomposition coefficient. Here, $i = 1, 2, \dots, l$ is the node number at 8th level of decomposition. N is the number of coefficients in each decomposed data. Energy ED_i , Standard deviation σ_i , mean μ_i , Kurtosis KUR_i , Skewness SK_i , Entropy EN_i , and Median med_i are computed for each node at 8th level of decomposition tree. The term $E(C_{ik} - \mu_i)^n$ is n th moment of the mean. Total seven properties, as defined in above equations, are used for creating the feature set. Thus at 8th level of decomposition, there are $256 (2^8 = 256)$ coefficients and a feature set of 1792 ($256 \text{ coefficients} \times 7 \text{ properties} = 1792$) features is extracted.

The total feature set becomes:

[Features] = [$ED_1, ED_2, \dots, ED_{256}, \sigma_1, \sigma_2, \dots, \sigma_{256}, \mu_1, \mu_2, \dots, \mu_{256}, KUR_1, KUR_2, \dots, KUR_{256}, SK_1, SK_2, \dots, SK_{256}, EN_1, EN_2, \dots, EN_{256}, med_1, med_2, \dots, med_{256}$].

The total numbers of nodes at 8th level of decomposition are 256. This matrix of length 1792×1 gives the total feature vector.

4. Proposed hybrid scheme using GA-SVM and wavelet transform

The proposed hybrid scheme based on wavelet packet decomposition, genetic algorithms and support vector machine is a novel scheme in terms of SVM acting as objective function for GA and the feature sub set selection by GA with stopping criterion set to 100% accurate classification rate. The present scheme starts with Wavelet Packet Decomposition of each signal up to 8th level. Two data sets namely A and E, having 100 EEG segments in each set, are taken for this experiment, as explained in Section 3.1. At 8th level of decomposition, each signal is decomposed into 256 coefficients. For each decomposition coefficient, seven properties namely energy, mean, median, standard deviation, kurtosis, skewness and Shannon entropy are calculated, giving $256 \times 7 = 1792$ features for each signal, making a complete feature set for a signal. These features are computed for both sets, i.e., set A and set E for creating a complete feature set for all the signals used in the experiment. In EEG signal classification, all these features are not important in the learning process of the classifier. Considering all the features for classification causes computational inefficiency and increases computation time. Sometimes, even the contrasting characteristics of some features may cause to degrade the overall performance of the classifier. So, for optimal classification accuracy and efficient computation time, optimal selection of features is carried out by a genetic algorithm with adaptive probability of mutation and SVM acting as objective function. The objective function values for GA are calculated by SVM. Out of the feature set comprising of 1792 features, the minimal feature set (50 features out of 1792) which gives best classification is selected optimally by the proposed GA-SVM scheme. The number of such features is of importance, as the exhaustive feature set (comprising of 1792 features) is computationally intensive and time consuming. A larger feature space than that used in [47] is chosen in the present study. A 6th level decomposition with lesser number of statistical and non-statistical properties and without optimal features selection scheme is presented in [47]. However, in the present work, more possibilities are explored to find optimal and relevant features for higher correct classification rate. Here optimal feature sub set means a feature sub set which will result

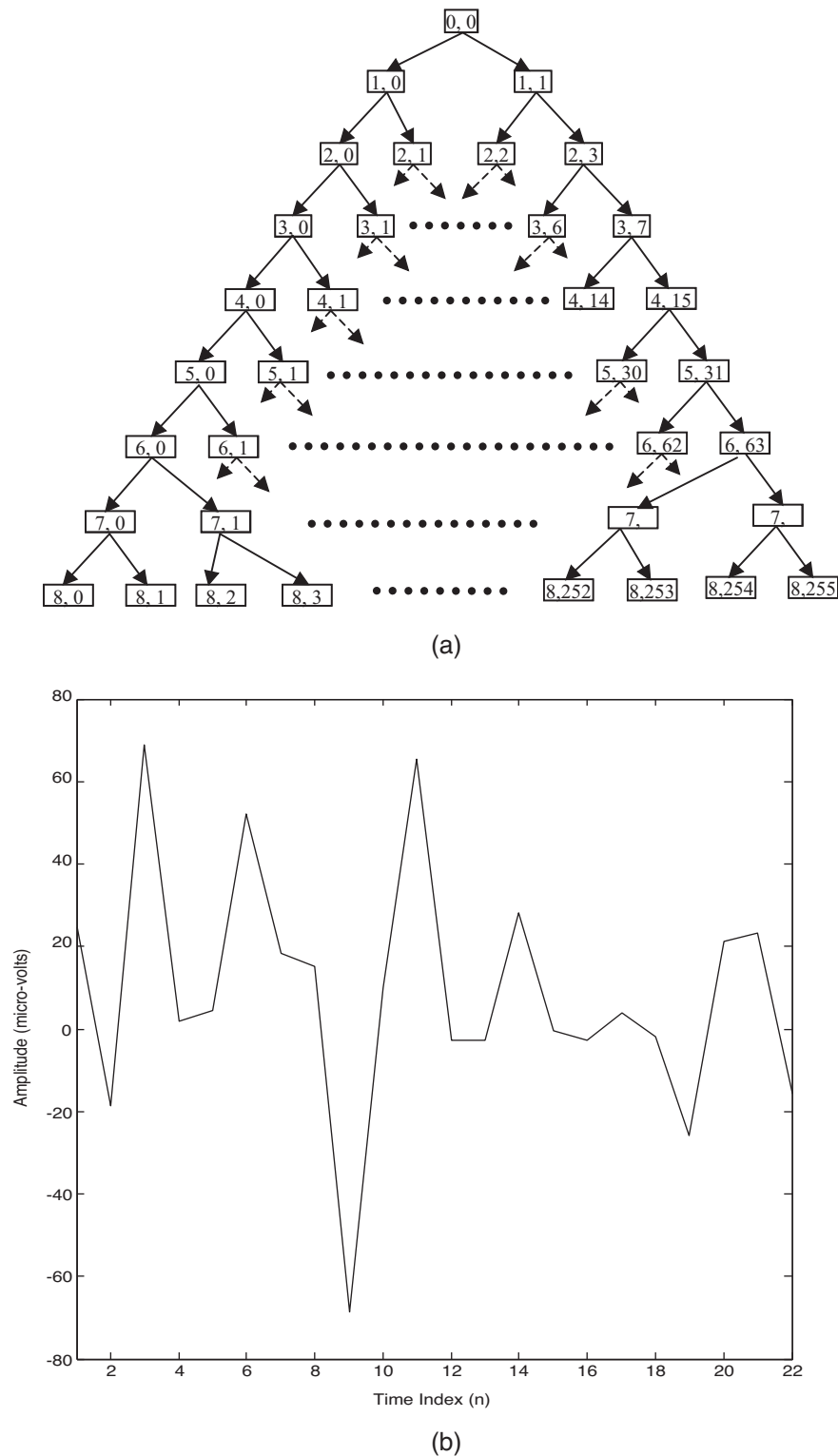


Fig. 2. (a) Wavelet packet analysis decomposition tree up to 8th level decomposition. (b) Data at node (8, 64).

in optimal (here maximum) classification accuracy. Fifty features out of the total feature set are selected. The number of optimal features is also an area of research, as to how many features will be sufficient for efficient and proper classification. According to the theory of large numbers, 10 is usually considered as the smallest large number, so authors started with selecting a features sub set of order 10 but this was insufficient to drive the proposed system to convergence, so the number of features increased in the steps of

10 towards the optimal features sub set and it is found that when 50 features were selected, then the system converged in reasonable time. As the number of features selected goes below 50, the number of features becomes insufficient for classification and the algorithm iterates in a futile way to search for a feature set which may give 100% classification rate. The stopping criterion is set to 100% classification rate, so the GA will stop only when it searches for a feature sub set which gives 100% classification rate. The proposed

technique is tested only for the present data set, however, if we put the technique for a clinical trial and new samples at large are presented before the proposed system, it may or may not result in 100% correct classification rate. The efficacy of the proposed scheme is further validated by testing it with a 10 fold cross validation scheme used for SVM with the selected feature sub set (50 features) for same data sets and it results in 100% correct classification rate with same order of specificity and sensitivity.

4.1. Genetic algorithm for feature selection

Genetic algorithm find optimal solution based upon the natural phenomenon of selection, reproduction and mutation. Fundamental work of Holland [21,22], Goldberg [11], Rechenberg [39], Schewefel [41], and Fogel [14] paved the way for today's genetic algorithms. When applied for practical problem solving, a GA begins with a few contending trial solutions. Variation operators are applied, and evaluation of objective function followed by a checking of stopping criterion completes the flow of GA, as shown in Fig. 3.

In present application, we have used GA with roulette wheel selection, one point cross over, mutation with adaptive probability of mutation and elitism. An objective measure of performance is used to assess the fitness of each trial solution and the selection mechanism determines which solutions should be maintained as 'parents' for the subsequent generation. A population of 30 chromosomes, with 50 allele positions in each chromosome, is initialized. Each allele value represents the index of a feature in the features set. The support vector machine is used for generating the fitness value for each solution (chromosome). Each chromosome

in the population is presented to the SVM and classification rate is calculated. The classification rate is the fitness value for the chromosome, and the objective is to maximize the classification rate. Roulette wheel selection is used for selection operator. One point crossover, with 0.8 probability of crossover is applied to perform crossover. The mutation rate is initially set as 0.009. Adaptive probability of mutation is used; as the GA reaches towards convergence, the mutation rate is decreased based upon the increase in average fitness of the population. Elitism is used to save so-far-best solutions during evolution of the GA, the best solution in each generation is saved and passed on to the next generation without applying variation operators to it. The fitness for the population is calculated again and stopping criterion is checked. The stopping criterion used in this paper is 100% correct classification rate. Initially, the GA took too much time to converge but as the probability of mutation is adaptively decreased with increase in average fitness of the successive populations, it converged earlier.

4.2. Support vector machine

The architecture of an SVM [52] is shown in Fig. 4. For the data sets $\{(x_1, y_1), \dots, (x_m, y_m)\}$, x are patterns of the data and y are the class labels of each data set x . In present application, there are two labels, epileptic and non-epileptic and are designated as +1 and -1. During training, the classes are presented with each data sample. The SVM tries to construct the optimal hyperplane yielding the maximum margin of separation between the members of the two classes by solving the following quadratic programming optimization problem. A soft margin can be constructed by introducing a

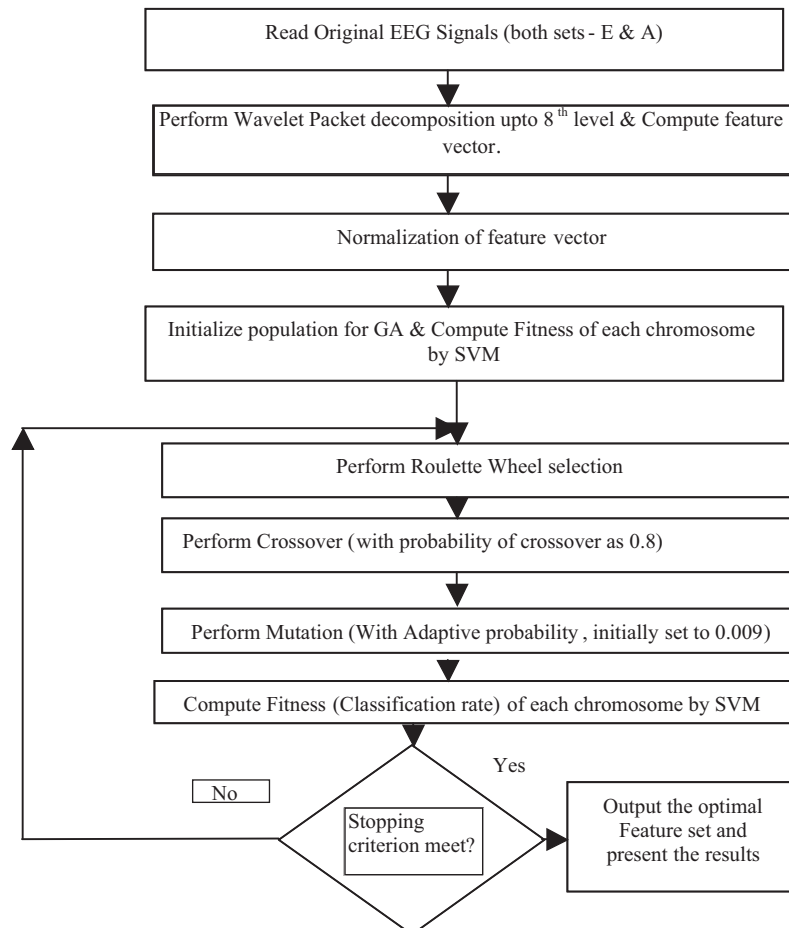


Fig. 3. Flow chart of GA-SVM classifier for epileptic seizure detection.

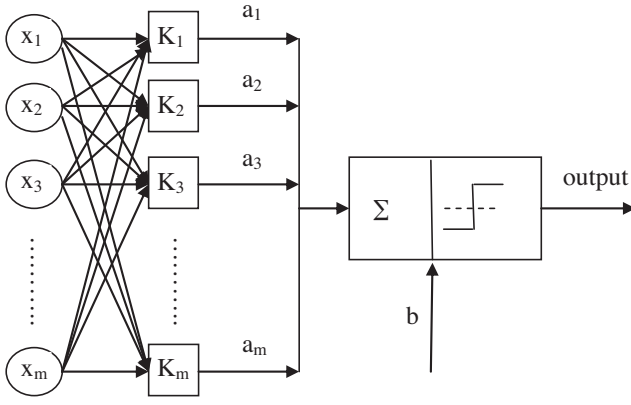


Fig. 4. Support vector machine.

slack variable β_i , $i = 1, \dots, m$, here m is the number of samples. The optimization problem becomes [52]:

$$\begin{aligned} &\text{Minimize}_{(W, b, \beta_i)} \quad \frac{1}{2} \|W\|^2 + D \sum_{i=1}^m \beta_i \\ &\text{Subject to} \quad y_i((W * X_i) + b) \geq 1 - \beta_i, \quad D > 0 \end{aligned} \quad (8)$$

$$\beta_i \geq 0, \quad i = 1, 2, \dots, m$$

$$W \in R^N, \quad b \in \mathbb{R}.$$

D is a user-defined constant greater than zero. This problem can be simplified by its Lagrangian dual as follows:

$$\begin{aligned} &\text{Minimize}_{(a \in \mathbb{R}^n)} \quad \sum_{i=1}^m a_i a_j y_i y_j (X_i X_j) \\ &\text{With} \quad D \geq a_i \geq 0 \\ &\sum_{i=1}^m a_i y_i = 0 \end{aligned} \quad (9)$$

$$\text{Here } W = \sum_{i=1}^m a_i y_i X_i$$

$$i = 1, 2, \dots, m.$$

$$\text{The decision function becomes } f(x) = \sum_{i=1}^m (a_i y_i (X, X_i) + b).$$

From this deduction, the patterns X_i ($a_i \neq 0$) are the support vectors. These support vectors lie on the margin of the decision boundary and all other data points are irrelevant. The feature vector X only appears in the equations as dot products so, the decision functions can be written as:

$$f(x) = \sum_{i=1}^m (a_i y_i K(X, X_i) + b).$$

where $K(X, X_i)$ is kernel function and is simply the dot product of X and X_i . This kernel function is used in the present application. Other types of kernel functions can be tried for making SVM more efficient and less time consuming for the proposed kind of heavy computation problem where the SVM is called many times in a single run of the GA.

4.3. Cross validation

Cross-validation, also called rotation estimation, is the statistical method of partitioning data set into subsets such that the analysis is initially performed on a single subset, while the other subsets are

retained for subsequent use in confirming and validating the initial analysis [9,10,29]. The initial subset of data is called the training set; the other subsets are called testing or validation sets. The original data set is partitioned into K subsets in a K -fold cross-validation arrangement. Out of these K subsets, one subset is kept as the validation data for testing the system, and the remaining $(K - 1)$ subsets are used as training data. The cross-validation is then repeated K times (equal to the no. of folds), with each of the K subsets used exactly once as the validation data. The K results from the folds then can be averaged to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation once only. The variance of the resulting estimate is reduced as K is increased. The disadvantage of this method is that the training algorithm has to be re-run from scratch K times, which means it takes K times as much computation to make an evaluation [29]. K -fold cross validation is not used for each computation of objective function, but only for the optimized set of features after the GA gives the final subset of features. Ten-fold cross validation is used in the present application.

Table 1
Optimally selected features by GA-SVM.

S. No.	Node No.	Feature index	Selected feature
1	10	67	Skewness
2	249	1742	Standard deviation
3	128	894	Mean
4	21	143	Kurtosis
5	114	797	Standard deviation
6	202	1413	Standard deviation
7	102	712	Mean
8	167	1166	Mean
9	214	1496	Mean
10	33	228	Skewness
11	119	831	Mean
12	200	1394	Energy
13	2	12	Mean
14	207	1446	Skewness
15	147	1026	Skewness
16	228	1596	Median
17	49	340	Skewness
18	7	44	Entropy
19	52	358	Energy
20	172	1202	Mean
21	91	634	Skewness
22	168	1070	Energy
23	135	940	Entropy
24	248	1731	Entropy
25	123	857	Kurtosis
26	57	396	Skewness
27	226	1580	Mean
28	128	893	Skewness
29	48	335	Standard deviation
30	206	1438	Kurtosis
31	227	1587	Mean
32	60	414	Energy
33	31	212	Entropy
34	68	474	Mean
35	162	1131	Skewness
36	187	1303	Energy
37	191	1333	Kurtosis
38	144	1005	Skewness
39	98	685	Standard deviation
40	250	1744	Energy
41	168	1174	Mean
42	136	952	Median
43	110	767	Skewness
44	19	130	Skewness
45	110	767	Skewness
46	250	1745	Entropy
47	245	1712	Skewness
48	174	1212	Energy
49	83	581	Median
50	17	114	Entropy

4.4. Performance evaluation

The following statistical measures [46] are calculated to evaluate the performance of the proposed GA-SVM model:

- (i) True positive (TP): The numbers of neural activities identified as seizure by the proposed system and also by medical expert.
- (ii) True negative (TN): The numbers of neural activities identified as normal by the proposed system and also by medical expert.
- (iii) False positive (FP): The numbers of neural activities identified as seizure by the proposed system and as normal by medical expert.
- (iv) False negative (FN): The numbers of neural activities identified as normal by the proposed system and as seizure by the medical expert.

Based upon the above statistical measures, the sensitivity, specificity and total accuracy of the system can be calculated [46] to evaluate the overall performance of the proposed system.

Sensitivity (SEN): The statistical performance of the classifier to detect epileptic signal.

$$SEN = TP / (TP + FN) \times 100\%.$$

Specificity (SPF): The statistical performance of the classifier to detect normal signal.

$$SPF = TN / (FP + TN) \times 100\%.$$

The accuracy is proportion of true results and given by:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100$$

5. Experimental results

The data sets are taken as explained in Section 3.1. The proposed scheme uses 8th level decomposition of data and complete features set is created as explained in Section 3.2. The total number of features is 1792. The genetic algorithm with support vector machine as objective function is used as optimization mechanism to select optimal features sub set, as explained in Section 4.1. The classification rate calculated by SVM acts as objective/fitness values for each chromosome. The best features subset of length 50 is selected by genetic algorithm with adaptive probability of mutation which was initially set at 0.009. The data are selected as 50% for training and 50% for testing, during the iterations of SVM-GA. Once optimal feature subset is selected by the GA-SVM system, the 10-cross validation approach is adopted. The 10-fold cross validation is implemented as explained in Section 4.3. The GA parameters are set as explained in Section 4.1. The present scheme gives 100% correct classification rate, hence found to be robust in epilepsy classification using EEG signals.

5.1. The selected feature subset

Out of the 1792 features, the GA optimally selected 50 features in order to improve classification rate and efficiency of the system. Here optimal selection of features means, the sub set of features for which GA optimizes the classification accuracy and meets its

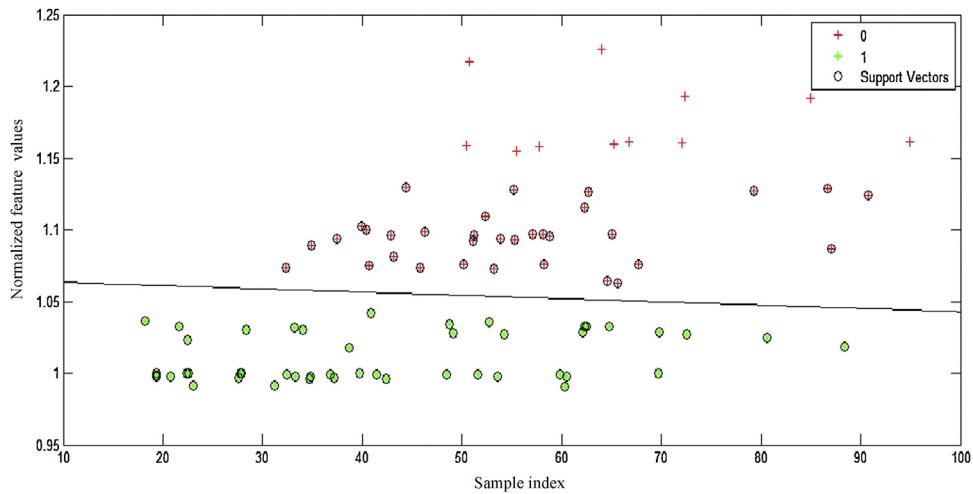


Fig. 5. Training of SVM.

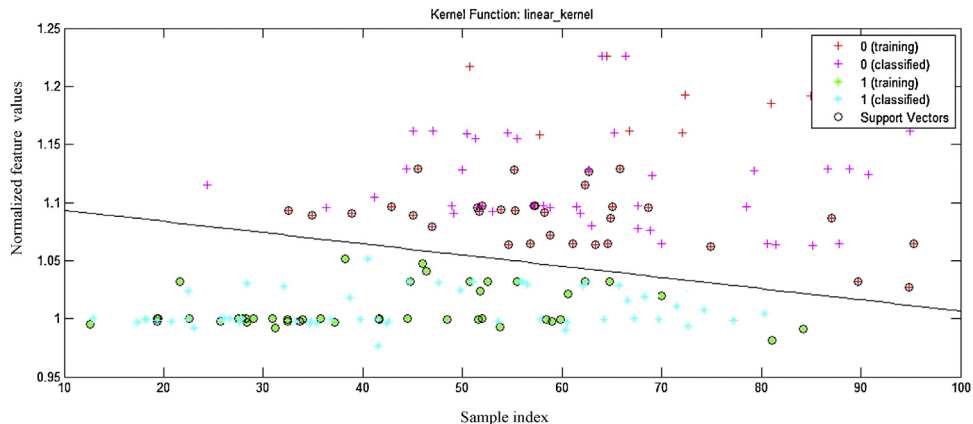


Fig. 6. Testing of SVM.

Table 2

Comparison of the proposed method with reported results in literature.

	Datasets used (A, B, C, D, E)	Train/test (%)	Sensitivity (%) [TP/(TP + FN) × 100]	Specificity (%) [TN/(FP + TN) × 100]	Total accuracy (%) [(TP + TN) × 100]/(TP + FP + FN + TN)
Kannathal et al. [26]	A, E	57/43	91.49	93.02	92.22
Guler et al. [16]	A, D, E	50/50	96.13	98.62	97.79
Mousavi et al. [32]	A, C, E	NA	91–96	91–96	91–96
Chandaka et al. [7]	A, E	50/50	92.00	100.00	95.96
Derya Ubeyli [11]	A, D, E	50/50	95.50	98.25	97.33
SVMA [47]	A, D, E	50/50	99.23	98.47	98.72
GA-SVM (Proposed)	A, E	50/50	100	100	100

stopping criterion of maximum classification accuracy i.e. 100% correct classification rate. If all 1792 features are presented to the classifier, it will be counterproductive in terms of computation resources and also some features are contradictory in terms of classification. Table 1 gives one such feature subset optimally selected by GA, which gives 100% classification rate. In this table, the first column gives Sr. No. and second column gives the node number. Here node number means the number of node at 8th level decomposition for which the corresponding feature has been calculated, node numbers are from 1 to 256 for present case because at 8th level of decomposition, there are 256 nodes for which coefficients are presented by the decomposition process. The feature index number gives the number of that feature in the total set of features (index numbers are from 1 to 1792). The feature set typically is structured as:

[Features] = $[ED_1, ED_2, \dots, ED_{256}, \sigma_1, \sigma_2, \dots, \sigma_{256}, \mu_1, \mu_2, \dots, \mu_{256}, KUR_1, KUR_2, \dots, KUR_{256}, SK_1, SK_2, \dots, SK_{256}, EN_1, EN_2, \dots, EN_{256}, med_1, med_2, \dots, med_{256}]$.

The features starting from ED_1 to med_{256} are numbered from 1 to 1792. The features set comprises of features calculated at all nodes of decomposition coefficients at 8th level of decomposition. The features are calculated by calculating the statistical and non-statistical properties at all nodes of decomposition tree at 8th level of decomposition. The last column gives the name of statistical/non-statistical property which is used to calculate the corresponding feature, e.g., the feature KUR_{256} , is obtained by calculating statistical property kurtosis for the coefficient at 256th node of 8th level decomposition.

5.2. Classification using SVM

During the execution of the GA-SVM programme, the classification rate is calculated by SVM for each chromosome of each population. Once the GA-SVM converges it gives final features set for which classification rate is 100%. As the stopping criterion for GV-SVM is set as 100% classification rate, so the optimal feature set containing 50 features gives 100% classification rate of detection of epilepsy. The training and testing graphs are respectively shown in Figs. 5 and 6 for the optimal feature set selected by the proposed GA-SVM.

5.3. Comparison with published results

The comparison of previously available techniques of epileptic seizure detection [52] in terms of classification accuracy is given in Table 2. Comparison of the proposed method (GA-SVM) is included in this table in last row along with other published methods as in [8,12,17,28,34,52]. Further, all these techniques have been designed and tested using the same kind of data. Individual datasets used by each study, the partition of the datasets between training/testing and the relevant performance measures are also enlisted in this table.

6. Conclusions and future directions

An automated system has been developed for detection of epileptic seizures from EEG signals. A new and robust method has been adopted for selecting optimal features by genetic algorithms to maximize the classification accuracy. The SVM acts as classifier which generates objective function values for the GA. SVM with linear kernel function is used in the present application. Other types of kernel function can be tried for making SVM more efficient and less time consuming for such kind of heavy computation problems where the SVM is called many times in a single run of the GA. For training of the system, the features calculated by statistical and non statistical properties of coefficients from nodes of the decomposition tree have been taken. The GA-SVM used for the epileptic seizure detection has been trained, cross validated and tested with the features extracted from Wavelet Packet decomposition of the EEG signals obtained from both normal as well as epileptic subjects. Out of 1792, only 50 relevant features are selected optimally by the GA-SVM system. The accuracy rate of 100% is achieved for detection by this robust classifier system, with 100% sensitivity and specificity for the data under consideration. The proposed technique is tested only for the present data set, however, if we put the technique for a clinical trial and new samples at large are presented before the proposed system, it may or may not result in 100% correct classification rate. The proposed technique outperforms the results of recent approaches reported in [20,37,52]. In present approach optimal selection of feature sub set is employed, so it outperforms the approaches of [47] in terms of feature set and also in terms of correct classification rate. The proposed approach outperforms the reported techniques in that it gives 100% correct classification rate whereas earlier the approaches reported the same between 73% [37] and 99.33% [47]. While Table 1 reports a classification rate from 92.22 to 98.72 (recently reported in 2012) for the data used as used in our experiment.

The higher level of accuracy, which is highly desirable in medical systems, makes this system a perfect helping tool for neurosurgeons. This system can also act as a robust real-time diagnostic tool for the next level of medical personnel posted in remote areas where there is an acute shortage of trained neurologists. The future scopes include developing a dedicated hardware machine for the same application based on the software technique presented in this paper. The hardware description language with appropriate programmable chips like FPGA may be used for creating the dedicated embedded system for this kind of machine.

Appendix A.

Wavelet Transforms are more suitable for processing of the non-stationary signals than fast Fourier transform (FFT) and the short time Fourier transform (STFT). Wavelet packet decomposition is a generalization of Discrete Wavelet decomposition that offers a

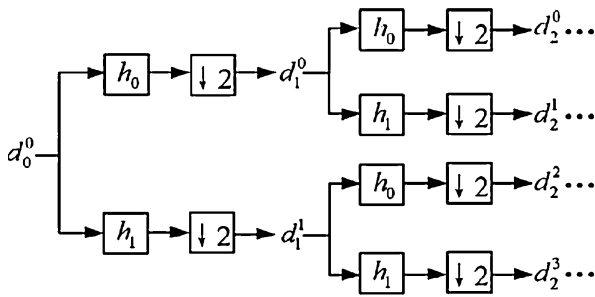


Fig. 7. Implementation of the wavelet packet transforms [50].

richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For a j -level decomposition, there are $j+1$ possible ways to decompose or encode the signal. In wavelet packet analysis, the details as well as the approximations are split. This yields more than 2^{2j-1} different ways to encode the signal (Fig. 7).

Wavelet packet transform decomposition algorithm of signal $f(t)$ is defined as [42,50]:

$$d_0^0(t) = f(t)$$

$$d_{j+1}^{2i}(t) = \sum_k h_0(k-2t)d_j^i, \quad i = 0, 1, \dots, 2^j - 1$$

$$d_{j+1}^{2i+1}(t) = \sum_k h_1(k-2t)d_j^i, \quad i = 0, 1, \dots, 2^j - 1$$

where $h_0(n)$ and $h_1(n)$ are a pair of orthogonal filters for decomposition: $h_0(n)$ is low pass filter and $h_1(n)$ is a high pass filter, in time domain these satisfy following equation:

$$h_1(k) = (-1)^k h_0(1-k)$$

Here, d_j^i represents the coefficient at i th node of j th level of decomposition.

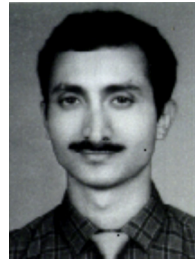
References

- [1] A.S. Zandi, et al., Predicting epileptic seizures in scalp EEG based on a variational bayesian Gaussian mixture model of zero-crossing intervals, *IEEE Trans. Biomed. Eng.* 60 (2013) 1401–1413.
- [2] A. Isaksson, et al., Computer analysis of EEG signals with parametric models, in: *Proceedings of the IEEE Conference*, 1981, pp. 451–461.
- [3] A. Petrosian, Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG, *Neurocomputing* 30 (2000) 201–218.
- [4] B. Litt, J. Echaz, Prediction of epileptic seizures, *Lancet Neurol.* 1 (2002) 22–30.
- [5] B. Litt, R. Esteller, J. Echaz, et al., Epileptic seizures may begin hours in advance of clinical onset: a report of five patients, *Neuron* 30 (2001) 51–64.
- [6] H. Berger, Über das Elektroenkephalogramm des Menschen, *Arch. Gen. Psychiat.* 87 (1929) 527–570.
- [7] C. Guerrero-Mosquera, et al., EEG feature selection using mutual information and support vector machine: a comparative analysis, in: *32nd Annual International Conference of the IEEE EMBS*, Buenos Aires, Argentina, August 31–September 4, 2010, pp. 4946–4949.
- [8] S. Chandaka, A. Chatterjee, S. Munshi, Cross-correlation aided support vector machine classifier for classification of EEG signals, *Expert Syst. Appl.* 36 (2) (2009) 1329–1336.
- [9] J. Chang, Y. Luo, K. Su, GPSM: a generalized probabilistic semantic model for ambiguity resolution, in: *Proceedings of the 30th Annual Meeting on Association for Computational Linguistics*, Newark, Delaware, June 28–July 02, Association for Computational Linguistics, Morristown, NJ, 1992, pp. 177–184.
- [10] P.A. Devijver, J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, London, 1982.
- [11] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
- [12] E. Derya Ubeyli, Statistics over features: EEG signals analysis, *Comput. Biol. Med.* 39 (8) (2009) 733–741.
- [13] E.D. Übeyli, Wavelet/mixture of experts network structure for EEG signals classification, *Expert Syst. Appl.* 34 (2008) 1954–1962.
- [14] L.J. Fogel, *Autonomous automata*, Ind. Res. 4 (1962) 14–19.
- [15] F. Mormann, R.G. Andrzejak, T. Kreuz, C. Rieke, P. David, C. Elger, K. Lehnertz, Automated detection of a pre-seizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients, *Phys. Rev. E* 67 (2003) 1–10.
- [16] G.R. Wu, F. Chen, D. Kang, X. Zhang, D. Marinazzo, H. Chen, Multiscale causal connectivity analysis by canonical correlation: theory and application to epileptic brain, *IEEE Trans. Biomed. Eng.* 58 (2011) 3088–3096.
- [17] N.F. Guler, E.D. Ubeyli, I. Guler, et al., Recurrent neural networks employing Lyapunov exponents for EEG signals classification, *Expert Syst. Appl.* 29 (2005) 506–514.
- [18] H. Feldwisch-Drentrup, et al., Statistical validation of event predictors: a comparative study based on the field of seizure prediction, *Phys. Rev. E* 83 (2011) 066704, <http://dx.doi.org/10.1103/PhysRevE.83.066704>.
- [19] H. Feldwisch-Drentrup, Identification of pre-seizure states in epilepsy: a data-driven approach for multichannel EEG recordings, *Front. Comput. Neurosci.* 5 (2011).
- [20] H. Ocak, Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy, *Expert Syst. Appl.* 36 (2009) 2027–2036.
- [21] J.H. Holland, *Adaptation in Natural & Artificial Systems*, The University of Michigan press, Ann Arbor, 1975.
- [22] J.H. Holland, Outline for a logical theory of adaptive systems, *J. Assoc. Comput. Mach.* 9 (3) (1962) 297–314.
- [23] Jerald Yoo, et al., An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor, *IEEE J. Solid-State Circuits* 48 (2013) 214–228.
- [24] J. Martinerie, C. Adam, M. LeVan Quyen, M. Baulac, S. Clemenceau, B. Renault, F. Varela, Epileptic seizures can be anticipated by non-linear analysis, *Nat. Med.* 4 (1998) 1173–1176.
- [25] K. Majumdar, Human scalp EEG processing: various soft computing approaches, *Appl. Soft Comput.* 11 (2011) 4433–4447.
- [26] K. Lehnertz, C. Elger, Can epileptic seizures be predicted? Evidence from non-linear time series analysis of brain electrical activity, *Phys. Rev. Lett.* 80 (1998) 5019–5022.
- [27] K. Sadatnezhad, et al., Classification of BMD and ADHD patients using their EEG signals, *Expert Syst. Appl.* 38 (2011) 1956–1963.
- [28] N. Kannathal, et al., Entropies for detection of epilepsy in EEG, *Comput. Methods Programs Biomed.* 80 (3) (2005) 187–194.
- [29] R. Kohavi, F. Provost, Glossary of terms. Editorial for the special issue on applications of machine learning and the knowledge discovery process, *Machine Learn.* 30 (1988) 271–274.
- [30] L. Iasemidis, J. Sackellares, H. Zaveri, W. Williams, Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures, *Brain Topogr.* 2 (1990) 187–201.
- [31] L. Guo, et al., Automatic feature extraction using genetic programming: an application to epileptic EEG classification, *Expert Syst. Appl.* 38 (2011) 10425–10436.
- [32] A. Liu, et al., Detection of neonatal seizures through computerized EEG analysis, *Electroencephalogr. Clin. Neurophysiol.* 82 (1992) 30–37.
- [33] M. Le Van Quyen, J. Martinerie, M. Baulac, F. Varela, Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings, *Neuroreport* 10 (1999) 2149–2155.
- [34] S.R. Mousavi, M. Niknazar, B.V. Vahdat, Epileptic seizure detection using AR model on EEG signals, in: *Cairo International Biomedical Engineering Conference (CIBEC 2008)*, 2008, pp. 1–4.
- [35] M. Heers, et al., Detection of epileptic spikes by magnetoencephalography and electroencephalography after sleep deprivation, *Seizure* 19 (2010) 397–403.
- [36] N. Robinson, et al., EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm, *IEEE Trans. Biomed. Eng.* 60 (2013) 2123–2132.
- [37] H. Ocak, Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm, *Signal Process.* 88 (2008) 1858–1867.
- [38] N. Pradhan, P.K. Sadasivan, G.R. Arunodaya, Detection of seizure activity in EEG by an artificial neural network: a preliminary study, *Comput. Biomed. Res.* 29 (1996) 303–313.
- [39] I. Rechenberg, *Cybernetic solution path of an experimental problem*, Royal Aircraft Establishment, Farnborough, Hants, UK, 1965, Library Translation Number 1122.
- [40] R.G. Andrzejak, et al., Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, *Phys. Rev. E* 64 (2001) 061907, <http://dx.doi.org/10.1103/PhysRevE.64.061907>.
- [41] H.P. Schwefel, *Numerical Optimization of Computer Models* (M. Finnis, Trans.), John Wiley, Chichester, 1977 (original work published 1977).
- [42] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, New York, 1999.
- [43] S.S. Viglione, G.O. Walsh, *Proceedings: epileptic seizure prediction*, *Electroencephalogr. Clin. Neurophysiol.* 39 (1975) 435–436.
- [44] B. Saraceno, G. Avanzini, P. Lee, Atlas: Epilepsy care in the world, WHO, Geneva, 2005 http://www.who.int/mental_health/neurology/epilepsy_atlas.r1.pdf

- [45] S. Gigola, F. Ortiz, C.E. D'Attellis, W. Silva, S. Kochen, Prediction of epileptic seizures using accumulated energy in a multiresolution framework, *J. Neurosci. Methods* 138 (1–2) (2004) 107–111.
- [46] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, *Expert Syst. Appl.* 32 (2007) 1084–1093.
- [47] T. Gandhi, et al., Expert model for detection of epileptic activity in EEG signature, *Expert Syst. Appl.* 37 (2010) 3513–3520.
- [48] U. Orhan, et al., EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, *Expert Syst. Appl.* 38 (2011) 13475–13481.
- [49] V. Navarro, J. Martinerie, M. Le Van Quyen, S. Clemenceau, M. Baulac, C. Adam, F. Varela, Seizures anticipation in human neocortical partial epilepsy, *Brain* 125 (2002) 640–655.
- [50] X. Jiao, et al., An algorithm for improving the coefficient accuracy of wavelet packet analysis, *Measurement* 47 (2014) 207–220.
- [51] X. Li, G. Ouyang, X. Yao, X. Guan, Dynamical characteristics of pre-epileptic seizures in rats with recurrence quantification analysis, *Phys. Lett. A* 333 (1–2) (2004) 164–171.
- [52] Y. Tang, D.M. Durand, A tunable support vector machine assembly classifier for epileptic seizure detection, *Expert Syst. Appl.* 39 (2012) 3925–3938.



Rohtash Dhiman completed his schooling from Government High School, Mohana, and Senior Secondary Certificate Examination from Hindu College, Sonipat, Haryana, India. He completed his Bachelor of Engg. (1998) and Master of Engg. (2008) degrees from C.R. State College of Engineering (Now D.C.R. University of Science and Technology) Murthal, Sonipat, India. Currently, he is Assistant Professor in Electrical Engineering Department at D.C.R. University of Science and Technology, Murthal, Sonipat, India, and pursuing Doctoral degree in Engg. from the same University. His current areas of interest include biomedical signal processing, intelligent techniques, brain computer interface, genetic algorithms, bio-inspired optimization techniques, machine vision and image processing.



J.S. Saini was born in Punjab, India, in 1961. He received the B.Tech. degree in Electrical Engineering from Punjab University, Chandigarh, India, in 1983 and the M.Tech. degree in Control Engineering and Instrumentation from the Indian Institute of Technology, New Delhi, India, in 1989. He was with the Electrical Engineering Department, Guru Nanak Engineering College, Ludhiana, India; Delhi Cloth Mills Engineering Products, Ropar, India; Punjab Engineering College, Chandigarh, India; Technical Teachers Training Institute, Chandigarh, India; and Regional Engineering College, Hamirpur, India, prior to joining the Electrical and Electronics Engineering Department, Chottu Ram State College of Engineering, (Now D.C.R. University of Science and Technology) Sonipat, India. Presently he is Professor in Electrical Engineering and Dean of Colleges in D.C.R. University of Science and Technology Murthal, Sonipat, India. He has published several research papers in refereed journals including IEEE Transactions & Conferences. His interests include chaotic systems, forecasting, fuzzy logic, and ANN applications in control and instrumentation. He was awarded the High Quality Presentation Award by the Systems Society of India for one of his research papers and Certificate of merit for another paper by Institution of Engineers (India).



Priyanka is working as Associate Professor at D.C.R. University of Science and Technology, Murthal, Sonipat, India. Her current areas of interest are signal processing, image processing and SAW filter design. Her highest qualification is Ph.D. in Electronics Engineering from Indian Institute of Technology, Delhi, India. She is IEEE member & has published several papers in refereed journals including IEEE Transactions & Conferences.